Improving the prediction of white matter hyperintensities evolution in brain MRI of patients with small vessel disease using stroke lesions information

Muhammad Febrian Rachmadi[a,c,**], Maria del C. Valdés-Hernández[b,*], Stephen Makin[d], Joanna Wardlaw[b] and Henrik Skibbe[a]

[a] *Brain Image Analysis Unit, RIKEN Center for Brain Science, Wako-shi, Japan*
[b] *Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom*
[c] *Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia*
[d] *University of Aberdeen, Aberdeen, United Kingdom*

ARTICLE INFO

Keywords:
deep prediction model
disease progression prediction
stroke lesions
white matter hyperintensities (WMH)

ABSTRACT

Predicting the evolution of white matter hyperintensities (WMH) (i.e., whether WMH will grow, remain stable, or shrink with time) is important for personalised therapeutic interventions. However, this task is difficult mainly due to the myriad of vascular risk factors (VRF) and comorbidities that influence the evolution of WMH, and the low specificity and sensitivity of the intensities and textures alone for predicting WMH evolution. Given the predominantly vascular nature of WMH, in this study, we evaluate the impact of incorporating stroke information to a probabilistic deep learning model to predict the evolution of WMH 1-year after the baseline image acquisition using brain T2-FLAIR MRI. The Probabilistic U-Net was chosen for this study due to its capability of simulating and quantifying uncertainties involved in the prediction of WMH evolution. We propose to use an additional loss called volume loss to train our model, and incorporate an influential factor of WMH evolution, namely, stroke lesions information. Our experiments showed that jointly segmenting the disease evolution map (DEM) of WMH and stroke lesions, improved the accuracy of the DEM representing WMH evolution. The combination of introducing the volume loss and joint segmentation of DEM of WMH and stroke lesions outperformed other model configurations with mean volumetric absolute error of 0.0092 *ml* (down from 1.7739 *ml*) and 0.47% improvement on average in shrinking, growing and stable WMH using Dice similarity coefficient.

1. Introduction

White matter hyperintensities (WMH) are one of the main neuroradiological features of cerebral small vessel disease (SVD) and have been commonly associated with stroke, aging, and dementia progression (Wardlaw et al., 2013; Prins and Scheltens, 2015; Wardlaw et al., 2017). They are often observed in T2-weighted and T2-fluid attenuated inversion recovery (T2-FLAIR) brain magnetic resonance images (MRI), appearing as bright regions. Small subcortical infarcts may be indistinguishable from WMH on structural MRI in absence of intravenous contrast due to sharing similar image intensity characteristics (Valdés Hernández et al., 2013), and if mistaken for WMH could negatively impact design of clinical research trials (Wang et al., 2012).

Clinical studies have indicated that some patients exhibit WMH progression over time (i.e., increasing in volume) (Schmidt et al., 2003; Sachdev et al., 2007; van Dijk et al., 2008) while some show WMH regression over time (i.e., shrinking in volume) (Moriya et al., 2009; Jochems et al., 2022a). Another study indicated that WMH dynamically change over time with clusters of WMH individually shrinking, staying unchanged (i.e., stable), or growing, these being observed at the same time point within the same individual (Ramirez et al., 2016). These variations have been associated with patients' comorbidities and clinical outcome (Chappell et al., 2017; Wardlaw et al., 2017). It must be noted that one clinical study also acknowledged that all factors influencing WMH evolution are still not fully known (Wardlaw et al., 2017; Jochems et al., 2022a).

Predicting the evolution of WMH is crucial for better care and prognosis of individual patients, but it remains a difficult task because of the different rate and direction of WMH evolution (Cai et al., 2022). Various deep learning models have been proposed to predict the evolution of WMH (Rachmadi et al., 2019, 2020, 2021). In these studies, WMH evolution is represented by a map called *disease evolution map* (DEM) which indicates the WMH voxels that shrink, grow, or remain stable at a further time point. DEM can be generated by subtracting images of manually labeled WMH from different time points. Previous studies generated the DEM by subtracting a baseline image of manually labeled WMH of a patient (Visit 1, V1) from a follow-up image of manually labeled WMH from the same patient one year after (Visit 2, V2) (Rachmadi et al., 2020, 2021). An example of DEM is visualised in Figure 1.

*Corresponding author
**Principal corresponding author
febrian.rachmadi@riken.jp (M.F. Rachmadi); M.Valdes-Hernan@ed.ac.uk (M.d.C. Valdés-Hernández)
ORCID(s): 0000-0003-1672-9149 (M.F. Rachmadi); 0000-0003-2771-6546 (M.d.C. Valdés-Hernández); 0000-0001-8701-9043 (S. Makin); 0000-0002-9812-6642 (J. Wardlaw); 0000-0002-2857-6023 (H. Skibbe)

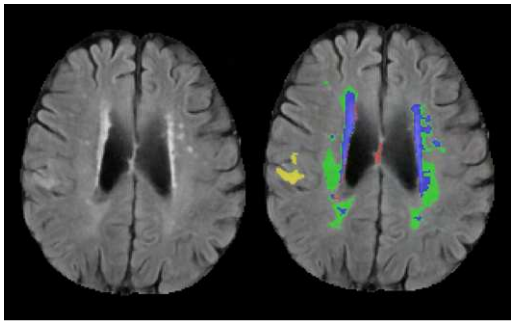

Figure 1: Visualisation of disease evolution map (DEM) of white matter hyperintesities (right) besides the corresponding brain-extracted FLAIR axial slice of the baseline scan or V1 (left). Red represents shrinking WMH, green represents growing WMH, and blue represents stable WMH. Note that labels of DEM for WMH evolution are combined with the label for the stroke lesions (yellow), which is one of the contributions of this study. Please, also observe that the growing WMH are in areas of ill-defined, subtle, slightly hyperintense regions on V1.

A recently proposed model for predicting the DEM of WMH based on a Probabilistic U-Net (Kohl et al., 2018), generates multiple predictions of DEM for a single brain MRI data (Rachmadi et al., 2021). This model was proposed to solve the challenge of representing spatial uncertainty (Rachmadi et al., 2020), given difficulties in distinguishing intensities and textures of shrinking and growing WMH in T2-FLAIR brain MRI. The uncertainty associated with the randomness of the dynamism of the WMH clusters is commonly known as *aleatoric uncertainty* (Hüllermeier and Waegeman, 2021). Previous experiments showed that models based on Probabilistic U-Net performed significantly better than the classical U-Net models previously used in predicting the evoultion of WMH using the DEM paradigm (Rachmadi et al., 2021).

So far, previous studies focused exclusively on the image modality as input and the appearance of WMH themselves while ignoring other clinically relevant factors. But a growing number of clinical studies have indicated that clinical factors such as previous strokes (Cai et al., 2022), age (van Dijk et al., 2008), and genetics (Schmidt et al., 2002, 2011; Godin et al., 2009; Luo et al., 2017), influence the rate and direction of WMH evolution. A previous study incorporated volume of stroke lesions as auxiliary input to the prediction model, but it did not improve the prediction results (Rachmadi et al., 2020). Thus, incorporating clinically associated factors into the model remains a challenge in the prediction of WMH evolution.

The **main contributions** of this study are two-fold:

1. **incorporating stroke lesions' information to the prediction model** (described in Section 3.2) and
2. **adding a volume loss to the cost function** (formulated as the mean squared error between the predicted and the reference WMH volumes, as per Equation 3 and described in Section 4.4.2) to improve prediction of WMH evolution.

We show that both significantly improve the prediction of DEM for WMH.

Other efforts related to this study are described in Section 2 while our proposed approaches are described in Section 3. All models tested in this study are based on the Probabilistic U-Net with adversarial training (Rachmadi et al., 2021) which is described in Section 3.1. Different configurations of all tested models are described in Section 3.3 and Table 1. All experimental settings are described in Section 4, and all results are discussed in Section 5. Lastly, our conclusion can be read in Section 6 while codes and trained models are available on GitHub [1].

2. Related Approaches

In general, previous studies that had developed prediction models for disease progression from medical image modalities using machine/deep learning can be categorised into the three different approaches listed below.

1. **Approaches predicting the outcomes of a disease.** These approaches are commonly used for diseases with high rates of mortality and disability. Some examples are those predicting the outcomes of COVID-19 (De Souza et al., 2021), multiple sclerosis (Pinto et al., 2020), and traumatic brain injury (Chong et al., 2015; Pease et al., 2022).
2. **Approaches predicting the progression of a disease with regards to the pathological timeline and/or commonly associated disease markers.** These approaches are commonly used for diseases with multiple stages of development and which take time to progress, such as dementia and Alzheimer's Disease (AD), with mild cognitive impairment (MCI) being their transitional stage (Pellegrini et al., 2018). Some examples are predicting conversion of MCI patients to AD (Zhang et al., 2021), conversion of healthy individuals to MCI and AD (Nakagawa et al., 2020), and predicting the progression of multimodal AD markers (e.g., ventricular volume, cognitive scores, etc.) (Nguyen et al., 2020).
3. **Approaches predicting dynamic changes (evolution) of specific disease features.** These approaches model and predict spatial changes of specific disease features such as evolution of WMH, enlargement of ventricles, and brain atrophy. Other examples are predicting lung nodule progression of pulmonary tumour (Rafael-Palou et al., 2022), predicting dynamic change of brain structures from healthy individuals to MCI and AD patients (Sauty and Durrleman, 2022), and studies for predicting the evolution of WMH in brain images of stroke patients (Rachmadi et al., 2019, 2020, 2021)

[1] https://github.com/febrianrachmadi/probunet-gan-vie

The present study belongs to the third category, in which a predictive model is used to estimate spatial dynamic changes of the evolution of WMH identified on an MRI scan. This third category is the most challenging because of the complexity and resolution of the data/image being predicted. While approaches in the first and second categories predict classes which are the disease outcomes (e.g., survive, death), classes of disease stages (e.g., MCI, AD), or associated disease markers (e.g., age, cognitive scores) from medical imaging data, approaches in the third category predict the dynamic changes and evolution of disease's features (e.g., lesions) that appear on images throughout the entire image space.

3. Proposed Deep Learning Model

Uncertainties are unavoidable when predicting the progression of WMH. Previous studies showed that incorporating uncertainties into a deep learning model, either by incorporating Gaussian noise as auxiliary input (Rachmadi et al., 2020) or using a conditional variational autoencoder in the shape of a Probabilistic U-Net with adversarial training (Rachmadi et al., 2021), improved prediction results, thus justifying the use of a Probabilistic U-Net with adversarial training (described in Section 3.1) in the present study. Section 3.2 details the two improvements made to the previously published probabilistic scheme (Rachmadi et al., 2021), implemented in the configurations related in section 3.3.

3.1. Probabilistic U-Net with adversarial training

In this study, we use Probabilistic U-Net with adversarial training as proposed in (Rachmadi et al., 2021) to capture spatial uncertainties from the brain MRI. Spatial aleatoric uncertainty is the biggest challenge in predicting the evolution of WMH, due to differences between experts in WMH delineation (i.e., ground truth reliability issues), and difficulty of previous automatic schemes distinguishing textures and intensities of shrinking and growing WMH in the T2-FLAIR MRI sequence (Rachmadi et al., 2020). This uncertainty cannot be reduced by simply adding more training data (Hüllermeier and Waegeman, 2021). Thus, a Bayesian deep learning model named Probabilistic U-Net (Kohl et al., 2018) proposed for this purpose, generated better prediction results than non-probabilistic models (Rachmadi et al., 2021).

The probabilistic U-Net with adversarial training consists of a U-Net configuration (Ronneberger et al., 2015), two variational encoders called Prior Net and Posterior Net, and a discriminator network for adversarial training. In this study, the U-Net was used as segmentation network for predicting the DEM. Prior Net and Posterior Net were used for variational inference. Prior Net estimates a low-dimensional Gaussian distribution called prior latent space by producing its mean(s) and variance(s) from T2-FLAIR MRI at baseline (i.e., V1, denoted x_{V1}). Whereas, Posterior Net estimates another low-dimensional Gaussian distribution called posterior latent space by producing its mean(s) and variance(s) from the follow-up T2-FLAIR MRI (i.e., V2, denoted x_{V2}) and ground truth DEM (y_{DEM}). *Kullback-Leibler* divergence is used during training to make prior and posterior latent spaces similar. In training, a sample $\boldsymbol{z}_{post}$ is taken from the posterior latent space ($\boldsymbol{z}_{post} \sim \mathcal{N}\left(\boldsymbol{\mu}_{post}, \boldsymbol{\sigma}_{post}\right)$) and then broadcasted and concatenated to the segmentation network. Multiple predictions of DEM ($\hat{y}^1_{DEM}, \hat{y}^2_{DEM}, \cdots, \hat{y}^n_{DEM}$) can be generated by using multiple samples ($z^1_{prior}, z^2_{prior}, \cdots, z^n_{prior}$) from the prior latent space ($\boldsymbol{z}_{prior} \sim \mathcal{N}\left(\boldsymbol{\mu}_{prior}, \boldsymbol{\sigma}_{prior}\right)$). Lastly, a discriminator network is used for adversarial training to enforce anatomically realistic DEM with regards to the T2-FLAIR MRI at V1 and V2. The Probabilistic U-Net used in this study is illustrated schematically in Figure 2. Detailed illustrations of each segmentation network, Posterior/Prior Net, and discriminator network are shown in Figures 3, 4 (left), and 4 (right) respectively.

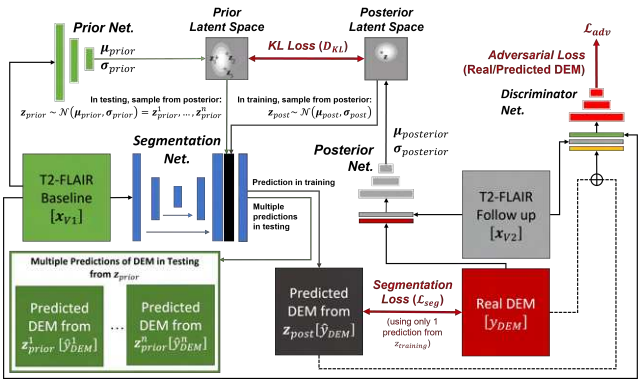

Figure 2: Schematic representation of the Probabilistic U-Net (Kohl et al., 2018) with adversarial training (Goodfellow et al., 2014) used in this study. This approach was first introduced in a previous work (Rachmadi et al., 2021).

3.2. Incorporation of stroke lesions information

Previous clinical studies have indicated that there are strong correlations between stroke occurrence and progression of WMH over time (Cai et al., 2022). In a previous study (Rachmadi et al., 2020), volume of stroke lesions was used as an auxiliary input to an scheme designed to estimate WMH evolution, but it was outperformed by the use of Gaussian noises as auxiliary input representing uncertainty. Thus, in this study, our main objective is to explore how information on stroke lesions can be incorporated to the probabilistic scheme, for better prediction of WMH evolution. We propose two different approaches which are 1) jointly segment the WMH disease evolution map (DEM) and stroke lesions, and 2) incorporating probabilistic maps of WMH change in relation to stroke lesions' locations.

3.2.1. Joint segmentation of DEM and stroke lesions

Due to the similar tissue intensity signal of WMH and ischaemic stroke lesions in T2-FLAIR brain MRI, we hypothesised that performing a joint segmentation of the WMH DEM and stroke lesions will improve the accuracy in the prediction of the WMH DEM because the deep learning

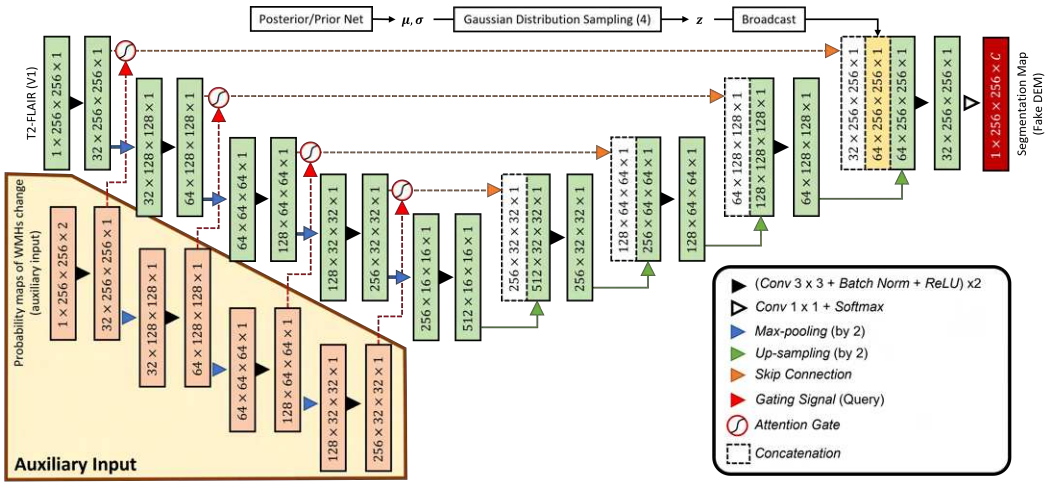

Figure 3: Segmentation network of Probabilistic U-Net used in this study, which is based on the original U-Net extended into Attention U-Net only when probability maps of WMH change are used as auxiliary input. The output channel of C is either 5 or 4 depending on whether stroke lesions are jointly segmented or not, respectively.

model will automatically learn the correlation between both features. In this approach stroke lesions do not need to be excluded in the preprocessing steps like in preceding works(Rachmadi et al., 2020, 2021). This newly proposed approach can be implemented by adding an output channel to the segmentation layer of the segmentation network originally with 4 channels (i.e., channels for background, shrinking WMH, growing WMH, and stable WMH), to 5 channels. Note in Figure 1 that the stroke lesions label has been added to the DEM of WMH.

3.2.2. Probabilistic maps of WMH change in relation to stroke lesions' locations

Results from a clinical study indicate that there are strong correlations between stroke lesions' location at baseline (V1) and WMH evolution after 1 year (V2) (Valdés Hernández et al., 2021) for patients with a stroke of type lacunar. Specifically, if stroke lesions are subcortical and located in either the *centrum semiovale* or the *lentiform nucleus* at V1, then there are significant changes to the WMH at V2 (both in volume and location) specific to the location of the stroke lesions at V1. This clinical study made available probability maps of WMH change indicating brain locations where changes of WMH are significant at V2 depending on the infarcted region after accounting for vascular risk factors (VRF) (Valdés Hernández et al., 2021).

We use these probability maps as auxiliary data input to an attention U-Net (Oktay et al., 2018) within the Probabilistic U-Net's segmentation network. In it, the information of the probability maps is encoded through the gating signal encoder (GSE), with outputs used as gating signal in multiple resolutions (see Figure 3). The general idea of this approach is to focus the attention of the segmentation network on the areas that have high probability of WMH change according to the locations of the stroke lesions.

Similar to the original Attention U-Net (Oktay et al., 2018), this study uses an additive attention gate (AG), but obtains the gating signals from the GSE instead of from the outputs of the next (coarser) convolutional block. The schematic of the additive AG can be seen in Figure 5. Input features (x_l) are from the U-Net's skip connections, gating signals (g_l) are from the gating signal encoder (GSE), α are the attention coefficients learned in the training process used to scale input features x_l to highlight important areas, $\bigoplus$ is an element-wise addition, $\bigotimes$ is an element-wise multiplication, and W_g, W_x, and ψ are $1\times1\times1$ convolution operations.

For implementing this approach, we perform brain parcellation and registration of the probability maps (in standard image space) to each patient's space to identify the locations of stroke lesions for each specific patient. Please see Section 4.2 for a detailed explanation of these processes. Note that this second proposed approach is more complex than the first proposed approach because it needs multiple preprocessing steps.

3.3. Configuration of the proposed approach

In this study we evaluate four configurations of segmentation networks. Three different configurations of networks incorporating probabilistic maps of WMH and/or stroke lesions were compared with the vanilla U-Net.

1. **PUNet**: Original U-Net (Ronneberger et al., 2015) was used for the segmentation network.
2. **PUNet-wSL**: Joint segmentation of DEM of WMH and stroke lesions was performed as explained in Section 3.2.1.
3. **Att-PUNet**: Attention U-Net with probabilistic maps of WMH change (as explained in Section 3.2.2) was used for segmentation network instead of the original U-Net.
4. **Att-PUNet-wSL**: Attention U-Net with probabilistic maps of WMH change was used for segmentation network and joint segmentation of DEM and stroke lesions was performed simultaneously.

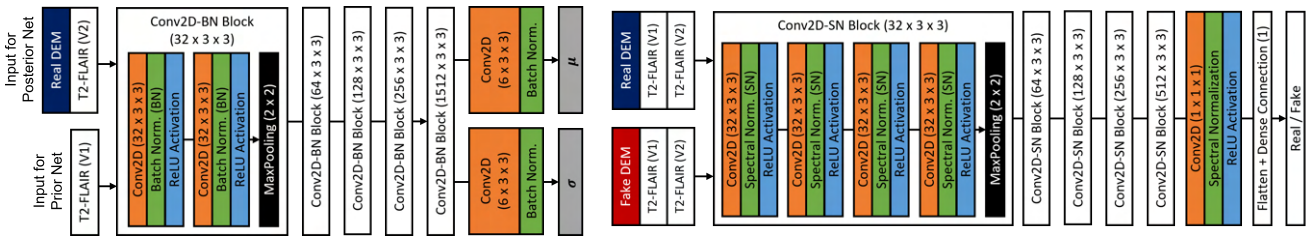

Figure 4: Architectures of Posterior/Prior Nets **(left)** and discriminator network **(right)** used in this study. Posterior/Prior Net. produces mean ($\boldsymbol{\mu}$) and standard deviation ($\boldsymbol{\sigma}$) that will be used to sample $\boldsymbol{z}$. Note that spectral normalization (Miyato et al., 2018) is used for the discriminator network.

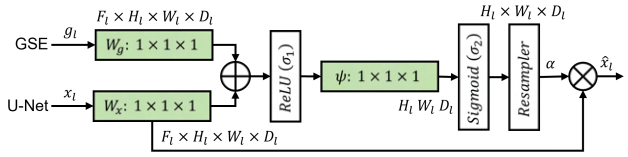

Figure 5: Schematic of additive attention gate (AG) used in this study and introduced in (Oktay et al., 2018). Input features (x_l) are from the U-Net's skip connection while gating signals (g_l) are from the gating signal encoder (GSE). Note that both x_l and g_l feature maps have the same size because they are from the same level of encoder (see Figure 3). Attention coefficients (α) are learned in the training process and used to scale input features x_l to highlight important areas.

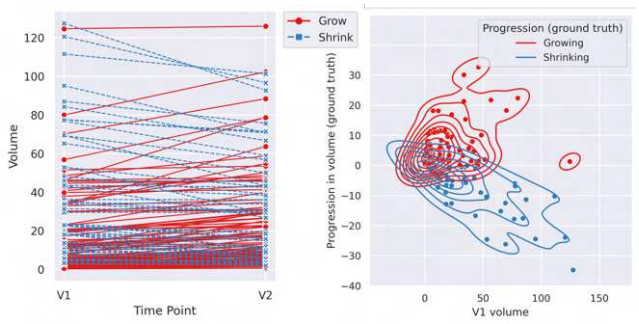

Figure 6: Volumetric progression of WMH (in *ml*) from V1 to V2 (1 year apart) for all subjects from our dataset **(left)** and distribution of volumetric progression of WMH (in *ml*) based on WMH volume at V1 for all subjects **(right)**.

4. Experimental Setting

This section describes the dataset (Section 4.1), training scheme (Section 4.3), and cost function (Section 4.4) used in this study.

4.1. Dataset

For comparability of our results with those previously published, we use the same dataset as (Rachmadi et al., 2020), which comprises MRI data from $n = 152$ patients that had a mild-to-moderate stroke and consented to participate in a study of stroke mechanisms (Wardlaw et al., 2017). The study protocols were approved by the Lothian Ethics of Medical Research Committee (REC 09/81101/54) and NHS Lothian R+D Office (2009/W/NEU/14), on the 29th of October 2009. All patients were imaged with the same acquisition protocol at two time points (i.e., baseline scan (V1), and a year after the baseline scan (V2)). In total, 304 MRI from 152 stroke patients (i.e., 152 V1 MRI and 152 V2 MRI) were used. Overall increase in WMH volume was identified in 98 of the 152 patients and reduction of WMH total volume in 54 patients. The magnitudes of WMH change (in *ml*) and their distribution for all patients can be seen in Figure 6.

All T2-FLAIR brain MRI were acquired with a GE 1.5T scanner, and a semi-automatic multi-spectral method was used to produce several brain masks including intracranial volume, cerebrospinal fluid, stroke lesions, and WMH, all which were visually checked and manually edited by an expert (Valdés Hernández et al., 2015). For the prediction of WMH evolution from V1 to V2, T2-FLAIR brain MRI at follow-up (V2) and T2-FLAIR brain MRI at baseline (V1) were linearly and rigidly aligned to a common space using FSL-FLIRT (Jenkinson et al., 2002). The space transformations were applied to all labels (i.e., binary/indexed masks) including manually-derived (i.e., after manually correcting results from a semi-automatic segmentation) labels of WMH. The spatial resolution of the images was $256 \times 256 \times 42$ with slice thickness of $0.9375 \times 0.9375 \times 4$ mm.

Also to facilitate comparability between methods and results, we used the same preprocessing pipeline as previous studies (Rachmadi et al., 2020, 2021). We generated a DEM for each patient by subtracting the manually corrected segmentation of WMH at V1 from the manually corrected segmentation of WMH at V2. Values of T2-FLAIR brain MRI were normalised into zero mean and unit variance for each patient. Data augmentations were performed, such as shifting, scaling, horizontal and vertical flip, and elastic transformation.

4.2. Data pre-processing for incorporation of probabilistic maps of WMH change

Given the influence of stroke lesion location in WMH change and evolution patterns when the stroke lesions are located at the *centrum semiovale* or the *lentiform nucleus* (Valdés Hernández et al., 2021), we only used probability

maps of WMH change based on stroke lesions incident at *centrum semiovale* or *lentiform nucleus*, publicly available[2].

Probability maps in the standard space were obtained from a clinical study (Valdés Hernández et al., 2021) and then registered to each patient's native space using niftyreg through TractoR (Clayden et al., 2011). To identify the location of stroke lesions within a human brain, an age-relevant brain template and its corresponding brain parcellation), also publicly available (Valdés Hernández, 2021), were registered to each patient's native space. If there were no stroke lesions at *centrum semiovale* or *lentiform nucleus* in a patient, then zero matrices were used as probabilistic maps. Both probabilistic maps for *centrum semiovale* or *lentiform nucleus* were concatenated before being used as auxiliary input in the segmentation network (see Figure 3 for illustration).

4.3. Training scheme

To make sure all patients are used in both training and testing, 4-fold cross validation with 512 epochs was performed with each fold consisting of 114 MRI for training and 38 for testing. The same 4-fold cross validation was used for our previous studies (Rachmadi et al., 2019, 2020, 2021). In the training phase, we used 14/114 MRI data for validation, and selected for testing the model that produced the best performance (i.e., lowest validation loss).

4.4. Cost function

We used three lost functions in training to optimize the different networks. These were: segmentation loss ($\mathcal{L}_{seg}$), probabilistic loss using *Kullback-Leibler* Divergence ($\mathcal{D}_{KL}$), and adversarial loss ($\mathcal{L}_{adv}$). We used the segmentation loss to compare the output of the segmentation network (i.e., the predicted DEM segmentation) against the ground truth of the DEM. The probabilistic loss was used to compare the similarity between prior and posterior latent spaces. And the adversarial loss was used to compare the similarity between the ground truth DEM and the predicted DEM.

4.4.1. Segmentation loss

For the segmentation loss, we used the weighted focal loss with γ (i.e., focal loss' hyperparameter) set to $\gamma = 2$ following the recommendation of the original paper (Lin et al., 2017). Equation 1 describes the weighted focal loss function for all pixels from an MRI slice where $y_{i,c} \in \{0, 1\}$ indicates the class membership for pixel i to class c, p_i the predicted probability that pixel i belongs to class c, and α_c is the weight for class c. The larger the value of α_c, the larger the contribution of class c to the loss value. P is the random variable for the predicted probability, Y is the random variable for the target classes, α are the weights for all classes, N is the number of pixels in an axial MRI slice (i.e., $N = 256$), and M is the number of classes in the DEM (i.e., $N = 4$ if stroke lesions are not automatically segmented and $N = 5$ if otherwise). Based on our preliminary experiments, the best weights were $\alpha_{c=0} = 0.25$ for background, $\alpha_{c=1} = 0.75$ for shrinking WMH, $\alpha_{c=2} = 0.75$ for growing WMH, $\alpha_{c=3} = 0.5$ for stable WMH, and $\alpha_{c=4} = 0.75$ for stroke lesions.

$$\mathrm{FL}(P, Y, \alpha) = -\sum_{i=1}^{N}\sum_{c=0}^{M=4} \alpha_c \, y_{i,c} \, (1 - p_{i,c})^{\gamma} \, log\left(p_{i,c}\right) \quad (1)$$

Note that the predicted segmentation of the DEM produced by the Probabilistic U-Net is conditioned to either the posterior or the prior latent space. In training, the predicted DEM segmentation is conditioned to the posterior latent space defined by $\boldsymbol{z}_{post} \sim \mathcal{N}\left(\boldsymbol{\mu}_{post}, \boldsymbol{\sigma}_{post}\right)$ and modelled by the Posterior Net. On the other hand, the predicted DEM segmentation is conditioned by the prior latent space that is formulated as $\boldsymbol{z}_{prior} \sim \mathcal{N}\left(\boldsymbol{\mu}_{prior}, \boldsymbol{\sigma}_{prior}\right)$ and modelled by the Prior Net in testing/inference. Thus, the probabilistic segmentation loss $\mathcal{L}_{seg}$ can be formulated as Equation 2 where $\hat{Y}_{DEM}$ is the predicted DEM segmentation.

$$\mathcal{L}_{seg} = \mathrm{FL}(P(\hat{Y}_{DEM}|X_{V1}, \boldsymbol{z}_{post}), Y, \alpha) + \mathrm{vol}(\hat{Y}_{DEM}, Y_{DEM}) \quad (2)$$

4.4.2. Volume loss

To avoid over- and under-segmentation in relation to the volume of WMH, a volume-loss (that is formulated as Equation 3) is added to Equation 2 as regularization term. The term keeps the volume of WMH from the predicted DEM ($\hat{Y}_{DEM}$) close to the volume of the WMH from the ground truth DEM (Y_{DEM}). To enforce this, we used the mean squared error (MSE). Note that only the classes $c = 2$ for growing WMH, and $c = 3$ for the stable WMH were used to calculate the volume of WMH. A denominator of 1000 was used to estimate the volume of WMH in *ml* (i.e., as voxel dimensions are in mm^3). To see the effectiveness of the volume-loss regularization, ablation studies were performed where the volume-loss was not used in training process for all configurations of the segmentation network explained in Section 3.3. All ablation studies and configurations are listed in Table 1.

$$\mathrm{vol}(\hat{Y}_{DEM}, Y_{DEM}) = \mathrm{MSE}\left(\frac{\sum_{c=2}^{M=3} \hat{y}_c}{1000}, \frac{\sum_{c=2}^{M=3} y_c}{1000}\right) \quad (3)$$

4.4.3. Probabilistic loss

We used *Kullback-Leibler* Divergence score ($\mathcal{D}_{KL}$) in the training process for training the Prior Net and Posterior Net. In this setting, Prior Net and Posterior Net were trained together with the Segmentation Net for predicting the DEM. Let Q be the posterior distribution from the Posterior Net and P be the prior distribution from the Prior Net. The difference between the posterior distribution Q and the prior distribution P is described by $\mathcal{D}_{KL}$ in Equation 4 where X_{V2} is the T2-FLAIR at V2, Y_{DEM} is the true DEM, and X_{V1} is

[2] https://datashare.ed.ac.uk/handle/10283/3934

Table 1
Names of tested models with their respective configurations. "PUNet" stands for Probabilistic UNet, "PMWCA" stands for Probabilistic Maps of WMH Change as Attention, and "JSDEMSL" stands for Joint Segmentation of DEM and Stroke Lesions.

Model's Name	PMWCA	JSDEMSL	Volume Loss
PUNet (Rachmadi et al., 2021)	-	-	-
PUNet-vol	-	-	✓
PUNet-wSL	-	✓	-
PUNet-wSL-vol	-	✓	✓
Att-PUNet	✓	-	-
Att-PUNet-vol	✓	-	✓
Att-PUNet-wSL	✓	✓	-
Att-PUNet-wSL-vol	✓	✓	✓

the T2-FLAIR at V1. Based on our preliminary experiments, the dimension for both $\boldsymbol{z}_{post}$ and $\boldsymbol{z}_{prior}$ is 4 (smaller than the original paper (Kohl et al., 2018) which used 6).

$$\mathcal{D}_{KL}(Q \parallel P) = \mathbb{E}_{\boldsymbol{z}_{post}\sim Q, \boldsymbol{z}_{prior}\sim P}[\log Q(X_{V2}, Y_{DEM}) - \log P(X_{V1})] \tag{4}$$

4.4.4. Adversarial loss

Similar to a previous study (Rachmadi et al., 2021), the original adversarial loss proposed by (Goodfellow et al., 2014) was slightly modified by adding a segmentation loss ($\mathcal{L}_{seg}$) so that the Segmentation Net was also optimised to produce better segmentation result. Similar to the original paper Goodfellow et al. (2014), the Segmentation Net aims at minimising Equation 5 while the discriminator network aims at maximising it. In Equation 5, G is the Segmentation Net, D is the discriminator network, $y \sim (X_{V1}, X_{V2}, Y_{DEM})$ is the joint distribution of T2-FLAIR MRI at V1 and V2 and ground truth DEM (i.e., X_{V1}, X_{V2}, and Y_{DEM} respectively), $x \sim (X_{V1}, X_{V2}, \hat{Y}_{DEM})$ is the joint distribution of T2-FLAIR MRI at V1 and V2 and predicted DEM (i.e., X_{V1}, X_{V2}, and $\hat{Y}_{DEM}$ respectively), $\mathbb{E}_y \sim Y_{GAN}$ is the expected value over Y_{GAN}, and $\mathbb{E}_x$ is the expected value over X_{GAN}.

$$\mathcal{L}_{adv} = \mathbb{E}_{y\sim Y_{GAN}}\left[\log(D(y))\right] + \mathbb{E}_{x\sim X_{GAN}}[\log(1-D(G(x))) + \mathcal{L}_{seg}(G(x))] \tag{5}$$

4.5. Evaluation measurements

In this study, we used the following evaluation measurements to assess the performance of all configurations listed in Table 1.

1. **Spatial agreement** between predicted and ground truth DEM is measured by the Dice similarity coefficient (DSC) (Dice, 1945), precision (PRE), and recall (REC). Higher values of DSC, PRE, and REC mean better performance. DSC, PRE, and REC can be calculated by using Equations 6, 7, and 8, respectively, where TP is true positive, FP is false positive and FN is false negative.

$$\text{DSC} = \frac{2 \times TP}{FP + 2 \times TP + FN} \tag{6}$$

$$\text{PRE} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{REC} = \frac{TP}{TP + FN} \tag{8}$$

2. **Uncertainty quantification and correlation analysis** to measure correlation between uncertainty values in predicted DEM and DSC values, is calculated as the Cross Entropy (CE) between the mean sample and all samples as per Equation 9 where γ is the uncertainty map, s is a set of predictions from an input, $\hat{s}$ is the mean sample of set s, CE is the cross entropy function, and $\mathbb{E}$ is the expected value function.

$$\gamma(\text{s}) = \mathbb{E}[\text{CE}(\hat{\text{s}}, \text{s})] \tag{9}$$

3. **Accuracy of prediction** assesses how good our proposed models predict WMH evolution for all patients (i.e., growing or shrinking). Accuracy of prediction for growing and shrinking WMH (i.e., subjects with growing and shrinking WMH are correctly predicted to have growing and shrinking WMH respectively) is calculated by the Equations 10 and 11 respectively. N_{GRW} and N_{SHR} are the number of subjects in our dataset who have growing and shrinking WMH. Whereas, P_{GRW} and P_{SHR} are the number of subjects correctly predicted as having growing and shrinking WMH.

$$\text{GRW} = \frac{P_{\text{GRW}}}{N_{GRW}} \tag{10}$$

$$\text{SHR} = \frac{P_{\text{GRW}}}{N_{\text{SHR}}} \tag{11}$$

4. **Estimated volume interval (EVI)** measures the deviation of the predicted WMH volume at follow-up (V2) from the lowest and highest possible predicted volumes of WMH (Rachmadi et al., 2021). The lowest and highest possible predicted volumes of WMH at V2 are estimated by ignoring the prediction channel for growing WMH and shrinking WMH respectively. In other words, the lowest possible volume of WMH (dubbed as Minimum Volume Estimation or 'MinVE') is assumed to occur when there are no growing WMH in the patient's brain. Whereas, the highest possible volume of WMH (dubbed as

Maximum Volume Estimation or 'MaxVE') is assumed to occur when there are no shrinking WMH in the patient's brain. There are 3 metrics in this evaluation: "CP" which stands for "Correct Prediction" (calculated by using Equation 12), "CPinEVI" which stands for "Correct Prediction in Estimated Volume Interval" (calculated by using Equation 13), and "(CP+WP)inEVI" which stands for "Correct Prediction + Wrong Prediction but still in EVI" (calculated by using Equation 14). In these equations, P^{in}_{GRW} and P^{in}_{SHR} are the number of subjects that are correctly predicted as having growing and shrinking WMH and have their estimated volumes of WMH at V2 are located between 'MinVE' and 'MaxVE'. Whereas, P^{in} is the number of subjects whose estimated volumes of WMH at V2 are located between 'MinVE' and 'MaxVE'.

$$\text{CP} = \frac{P_{\text{GRW}} + P_{\text{SHR}}}{N_{\text{GRW}} + N_{\text{SHR}}} \tag{12}$$

$$\text{CPinEVI} = \frac{P^{in}_{\text{GRW}} + P^{in}_{\text{SHR}}}{N_{\text{GRW}} + N_{\text{SHR}}} \tag{13}$$

$$\text{(CP+WP)inEVI} = \frac{P^{in}}{N_{\text{GRW}} + N_{\text{SHR}}} \tag{14}$$

5. **Volume error** measures how close the predicted WMH volumes are with the real WMH volumes at the follow-up assessment (V2). Volume error can be calculated by using Equation 15 where vol^{V2}_{true} is the true volume of WMH at V2, $\text{vol}^{V2}_{predicted}$ is the predicted volume of WMH at V2, and vol^{V2}_{error} is the volume error.

$$\text{vol}^{V2}_{error} = \text{vol}^{V2}_{predicted} - \text{vol}^{V2}_{true} \tag{15}$$

5. Results and Discussions

In this section, we show and discuss the results of the evaluations using the four performance measurements described in Section 4.5, namely spatial agreement evaluation (Section 5.1), qualitative/visual evaluation (Section 5.2), volume based evaluation (Section 5.3), and uncertainty quantification (Section 5.4), for all model configurations listed in Table 1.

5.1. Spatial agreement evaluation

Tables 2, 3, and 4 show performances of all model configurations listed in Table 1 evaluated using DSC, PRE, and REC, calculated using Equations 6, 7, and 8 respectively. The best and second best measurement values for each DEM label are written in bold and underlined respectively. Note that the label 'Changing' refers to shrinking and growing WMH combined together as one label. The 'Stroke Lesions' label is only available when joint segmentation of WMH DEM and stroke lesions are performed (see Section 3.2.1).

From Table 2, we can see that joint segmentation of DEM and stroke lesions with volume loss (PUNet-wSL-vol) produced the best segmentation results based on DSC for 'Shrinking' (0.2290) and 'Average' (0.3598). Furthermore, we can see that joint segmentation of DEM and stroke lesions (described in Section 3.2.1) by PUNet-wSL (i.e., without volume loss) and PUNet-wSL-vol (i.e., with volume loss) produced either the best or second best DSC values for all categories of DEM except for 'Growing' and 'Stable' WMH, which were achieved by the original configuration either without volume loss (PUNet) or with volume loss (PUNet-vol). On the other hand, other configurations especially with auxiliary input of probabilistic maps of WMH change, described in Section 3.2.2, (i.e., Att-PUNet, Att-PUNet-vol, Att-PUNet-wSL, and Att-PUNet-wSL-vol) failed to improve the performance of the DEM segmentation while improved the performance of 'Stroke Lesions' segmentation.

We can also see from Table 2 that models trained using volume loss (Equation 3, Section 4.4.2), which are PUNet-vol, Att-PUNet-vol, PUNet-wSL-vol, and Att-PUNet-wSL-vol, produced better DSC values on 'Average'. Note that 'Average' DSC value is calculated by averaging DSC values of 'Shrinking', 'Growing', and 'Stable'. This indicates that the volume loss impacted positively in the task of estimating the DEM of WMH.

Tables 3 and 4 show that joint segmentation of DEM and stroke lesions (i.e., PUNet-wSL and PUNet-wSL-vol), in general, also produced either the best or second best PRE and REC values for some categories while probabilistic maps of WMH change as auxiliary input (Att-PUNet) produced the best PRE values, and the original configuration (PUNet) produced the best REC values for most DEM categories. These results show that performing joint segmentation of DEM and stroke lesions as described in Section 3.2.1, either by using volume loss (PUNet-wSL-vol) or not (PUNet-wSL), improved the quality of the predicted DEM of WMH producing balanced performances in DSC, PRE, and REC metrics. FP and FN counts produced by other configurations were imbalanced. This imbalance in FP and FN counts influenced the DSC, PRE, and REC values through their respective equations (Equations 6, 7, and 8 respectively). For example, PUNet-vol produced high DSC and REC values and low PRE values due to the combination of higher FP and lower FN counts. Whereas, Att-PUNet produced high PRE values and low DSC and REC values due to the combination of lower FP and higher FN counts.

To provide a better illustration of the relationship between DSC, PRE, and REC values and FP and FN counts, we present the confusion matrices and a table compiling these values from the 'Shrinking' WMH and 'Growing' WMH labels obtained from PUNet-vol and PUNet-wSL-vol configurations (Figure 7 and Table 5 respectively). Figure 7 contains the number of segmented voxels corresponding

Table 2
Dice similarity coefficient (DSC) for all model configurations listed in Table 1. Symbol ↑ indicates that higher values are better. The best and second best measurement values for each category of WMH are written in bold and underlined respectively.

Model's Name	Dice Similarity Coefficient (DSC) ↑					
	Shrinking	Growing	Stable	Average	Changing	Stroke Lesions
PUNet (Rachmadi et al., 2021)	0.2132	0.2137	0.6385	0.3551	0.3633	-
PUNet-vol	0.2107	**0.2232**	**0.6439**	0.3593	0.3642	-
PUNet-wSL	0.2217	0.2130	0.6437	0.3595	**0.3719**	0.4499
PUNet-wSL-vol	**0.2290**	0.2112	0.6392	**0.3598**	0.3681	0.4281
Att-PUNet	0.2211	0.1796	0.6302	0.3437	0.3510	-
Att-PUNet-vol	0.2078	0.1981	0.6315	0.3458	0.3471	-
Att-PUNet-wSL	0.1968	0.2045	0.6240	0.3417	0.3543	0.5338
Att-PUNet-wSL-vol	0.1960	0.2077	0.6322	0.3453	0.3536	**0.5430**

Table 3
Precision (PRE) for all model configurations listed in Table 1. Symbol ↑ indicates that higher values are better. The best and second best measurement values for each category of WMH are written in bold and underline respectively.

Model's Name	Precision (PRE) ↑					
	Shrinking	Growing	Stable	Average	Changing	Stroke Lesions
PUNet (Rachmadi et al., 2021)	0.2349	0.2331	0.6638	0.3773	0.3489	-
PUNet-vol	**0.2527**	0.2391	0.6553	0.3824	0.3686	-
PUNet-wSL	0.2346	**0.2541**	0.6600	0.3829	0.3690	**0.4226**
PUNet-wSL-vol	0.2295	0.2479	0.6642	0.3805	0.3603	0.4065
Att-PUNet	0.2370	0.2435	**0.6740**	**0.3848**	**0.3694**	-
Att-PUNet-vol	0.2241	0.2253	0.6516	0.3670	0.3501	-
Att-PUNet-wSL	0.2245	0.2247	0.6385	0.3626	0.3461	0.3918
Att-PUNet-wSL-vol	0.2315	0.2345	0.6531	0.3730	0.3628	0.3746

to each label (n) from all patients in the testing set, false negative rate (fnr), false positive rate (fpr), true positive rate (TPR), and positive predictive value (PPV). Table 5 compiles values of DSC, PRE, REC, FN, and FP for the 'Shrinking' WMH and 'Growing' WMH labels from both PUNet-vol and PUNet-wSL-vol configurations. From both, Figure 7 and Table 5, we can see that PUNet-vol produced higher PRE value for 'Shrinking' WMH with lower FP counts than PUNet-wSL-vol. But PUNet-vol produced lower PRE value for 'Growing' WMH as it produced higher FP counts than PUNet-wSL-vol in this label/category.

Figure 8 shows the correspondence between the individual WMH volumes at V1 and the DSC values for all DEM labels produced by the four models that have the WMH volume loss incorporated. The ideal scenario will be a thin cloud of points aligned horizontally throughout the volume range, near and below the DSC value of 1 for all the labels. As can be appreciated all the four models produced similar results with DSC values around 0.8 for 'Stable' WMH volumes above 30 ml, while 'Shrinking' and 'Average' had DSC values between 0.3 and 0.5 and 'Growing' had the lowest DCS values across the same range of WMH volumes. It is worth noting that DSC values for 'Shrinking' and 'Stable'

Table 4
Recall (REC) for all model configurations listed in Table 1. Symbol ↑ indicates that higher values are better. The best and second best measurement values for each category of WMH are written in bold and underlined respectively.

Model's Name	Recall (REC) ↑					
	Shrinking	Growing	Stable	Average	Changing	Stroke Lesions
PUNet (Rachmadi et al., 2021)	0.2730	**0.2646**	0.6783	**0.4053**	**0.4091**	-
PUNet-vol	0.2408	0.2569	**0.6881**	0.3953	0.3820	-
PUNet-wSL	0.2979	0.2303	0.6814	0.4032	0.4012	0.3811
PUNet-wSL-vol	**0.3066**	0.2346	0.6703	0.4038	0.4032	0.3831
Att-PUNet	0.2885	0.1806	0.6450	0.3714	0.3565	-
Att-PUNet-vol	0.2579	0.2319	0.6785	0.3894	0.3797	-
Att-PUNet-wSL	0.2427	0.2413	0.6768	0.3869	0.3877	**0.4048**
Att-PUNet-wSL-vol	0.2267	0.2323	0.6726	0.3772	0.3710	0.3846

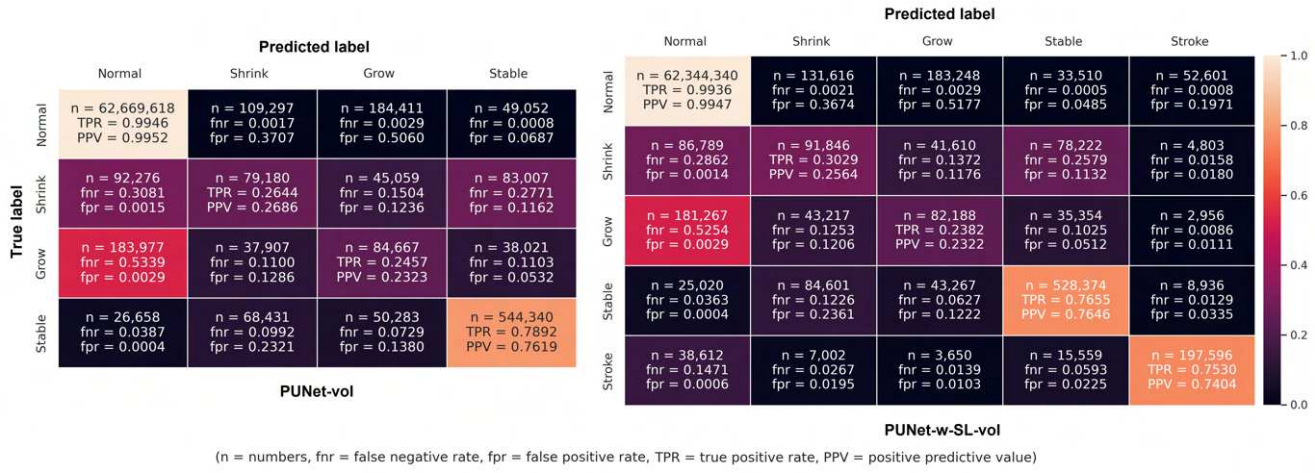

Figure 7: Confusion matrices for all labels produced by PUNet-vol and PUNet-wSL-vol configurations from all subjects. Abbreviation *n* stands for number of segmented voxels which can be used to calculate false negative rate (*fnr*), false positive rate (*fpr*), true positive rate (TPR), and positive predictive value (PPV). Note that TPR and *fnr* are calculated horizontally for each row (true label of DEM). On the other hand, PPV and *fpr* are calculated vertically for each column (predicted label of DEM).

Table 5
Comparison of DSC, PRE, and REC values to FN and FP counts for PUNet-vol and PUNet-wSL-vol configurations. Symbols ↑ and ↓ indicate that higher and lower values are better respectively.

	Shrinking WMH					Growing WMH				
	DSC ↑	**PRE ↑**	**REC ↑**	**FN ↓**	**FP ↓**	**DSC ↑**	**PRE ↑**	**REC ↑**	**FN ↓**	**FP ↓**
PUNet-vol	0.2107	**0.2527**	0.2408	220,342	**215,635**	**0.2232**	0.2391	**0.2569**	**259,905**	279,753
PUNet-wSL-vol	**0.2290**	0.2295	**0.3066**	**211,424**	266,436	0.2112	**0.2479**	0.2346	262,794	**271,775**

WMH (i.e., and, therefore 'Average' WMH) from patients with smaller WMH volumes were lower than those from patients with WMH volumes of 30 ml and above, bringing the overall mean DSC, represented by the horizontal line in the graphs, to a lower value. However, the mean DSC for 'Growing' WMH seem to be rather higher than most of the individual results, as this label was slightly better identified in the scans from patients that had low WMH volume.

Confusion matrices in Figure 7, show a high level of uncertainty between 'Growing' WMH and 'Normal' brain tissues as more than 50% of the 'Growing' WMH identified in the ground truth DEM were wrongly predicted as 'Normal' tissues (i.e., under-segmentation of 'Growing' WMH which leads to higher *fnr* in the confusion matrix) by PUNet-vol and PUNet-wSL-vol configurations with $fnr = 0.5339$ and $fnr = 0.5254$ respectively. In extended experiments, all proposed configurations were observed producing the same level of under-segmentation for 'Growing' WMH. In general, areas of 'Growing' WMH are difficult to be differentiated from 'Normal' brain tissues due to the high

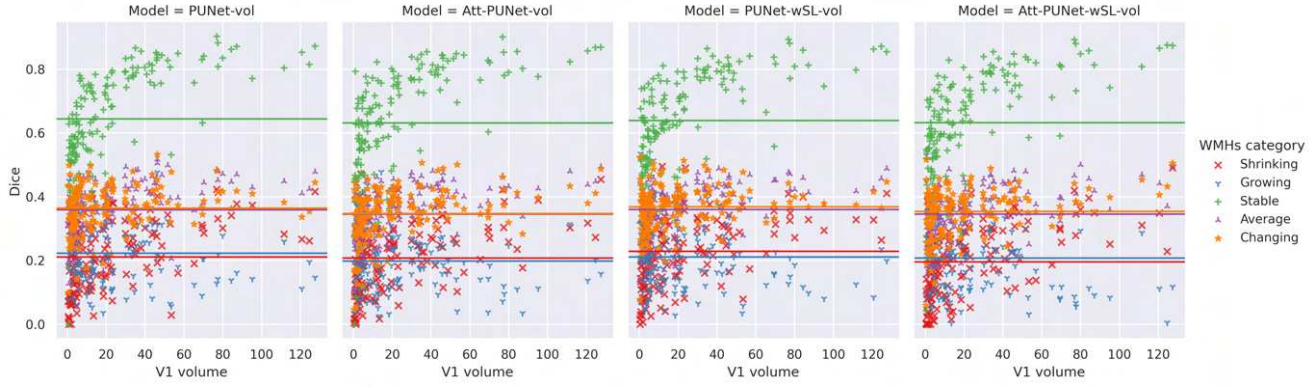

Figure 8: Correspondence between WMH volume (*ml*) at V1 and DSC values for all DEM labels produced by PUNet-vol, Att-PUNet-vol, PUNet-wSL-vol, and Att-PUNet-wSL-vol configurations.

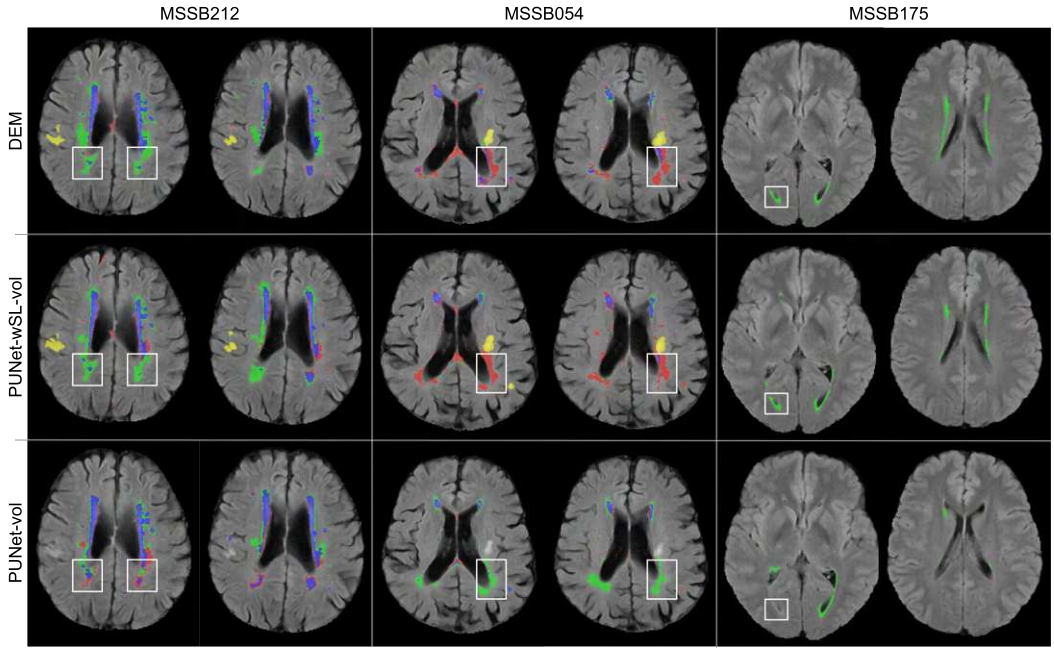

Figure 9: Examples of predicted DEM produced by PUNet-wSL-vol and PUNet-vol and their corresponding DEM ground truth from subjects with high DSC values on average. Red represents shrinking WMH, green represents growing WMH, blue represents stable WMH, and yellow represents stroke lesions. Obvious improvements are highlighted in white rectangles.

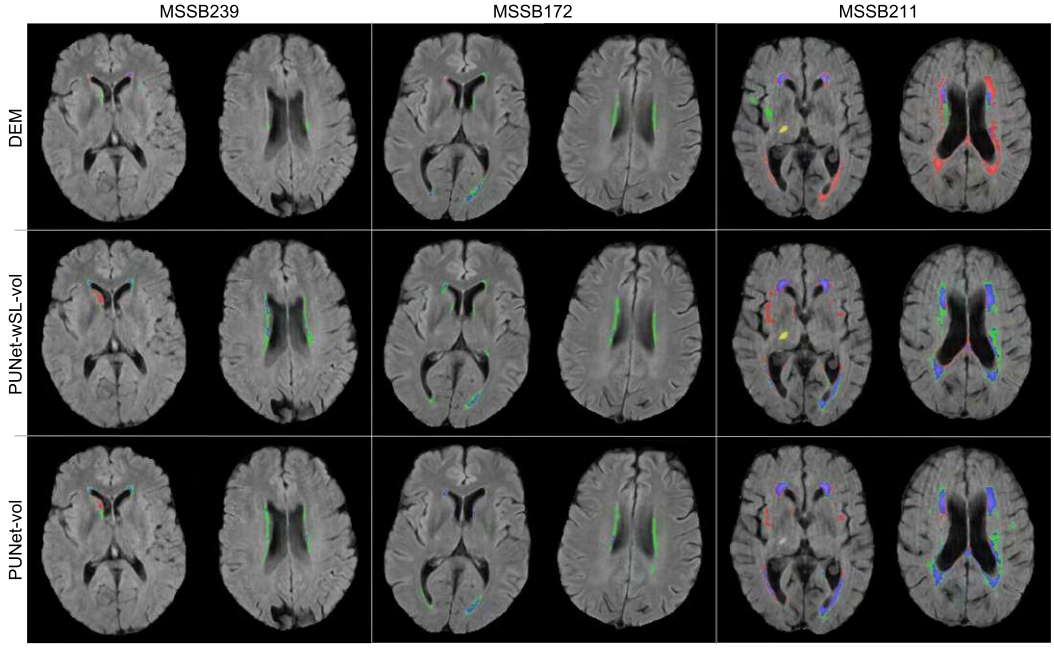

Figure 10: Examples of predicted DEM produced by PUNet-wSL-vol and PUNet-vol and their corresponding DEM ground truth from subjects with low DSC values on average. Red represents shrinking WMH, green represents growing WMH, blue represents stable WMH, and yellow represents stroke lesions.

level of uncertainty between these two classes. Overall, for the model that jointly segmented the stroke lesions and the WMH, mean DSC values were slightly better in this sample.

5.2. Qualitative/visual evaluation

PUNet-wSL-vol and PUNet-vol were chosen for qualitative/visual evaluation as they produced the best and second best DSC values on 'Average' (See Table 2 Section 5.1).

Figures 9 and 10 show examples of the predicted DEM segmentation from PUNet-wSL-vol and PUNet-vol and their corresponding DEM ground truth forpatients with high

Table 6
Volume based evaluation for all models evaluated. There are 98 patients with growing (GRW) and 54 with shrinking (SHR) volume of WMH. "CP" stands for "Correct Prediction", "CPinEVI" stands for "Correct Prediction in Estimated Volume Interval", and "(CP+WP)inEVI" stands for "Correct Prediction + Wrong Prediction but still in EVI". Symbol ↑ indicates that higher values are better while symbol → 0 indicates that values closer to 0 are better. The best and second best values for each evaluation measurements are written in bold and underlined respectively.

Model's Name	Prediction ↑		Estimated Volume Interval (n=152) ↑			Volumetric Error
	GRW	SHR	CP	CPinEVI	(CP+WP)inEVI	(std) → 0
PUNet (Rachmadi et al., 2021)	78.57%	46.30%	67.11%	47.37%	<u>61.18%</u>	-1.7739 (9.798)
PUNet-vol	<u>83.67%</u>	51.85%	71.71%	46.71%	60.53%	-0.8342 (8.657)
PUNet-wSL	75.51%	64.81%	71.71%	48.68%	59.21%	<u>0.2269 (10.427)</u>
PUNet-wSL-vol	74.49%	<u>74.07%</u>	<u>74.34%</u>	**53.29%**	**62.50%**	**-0.0092 (9.751)**
Att-PUNet	70.41%	**79.63%**	73.68%	45.39%	55.26%	3.1823 (8.447)
Att-PUNet-vol	81.63%	55.56%	72.37%	43.42%	54.61%	-0.5546 (9.043)
Att-PUNet-wSL	**86.73%**	55.56%	**75.66%**	<u>51.97%</u>	59.87%	-0.5978 (10.901)
Att-PUNet-wSL-vol	81.63%	64.81%	**75.66%**	43.42%	53.95%	0.2701 (9.050)

and low DSC values on 'Average' respectively. Figure 9 shows that PUNet-wSL-vol, which jointly segments WMH DEM and stroke lesions, produced better segmentation results than PUNet-vol, which exhibits a high level of uncertainty in predicting shrinking and growing WMH. Confusion matrices in Figure 7 show that PUNet-wSL-vol lowered this uncertainty by producing lower rates of *fnr* (and their corresponding FN counts (*n*)) for shrinking and growing WMH) in most cases. Figure 10 illustrates cases where low DSC values of predicted WMH DEM were caused mostly by two reasons: low WMH volume at V1 (patients MSSB239 and MSSB172) and brain MRI artefacts (patients MSSB239 and MSSB211). Based on our observations, these two problems were relevant throughout the sample in our evaluations.

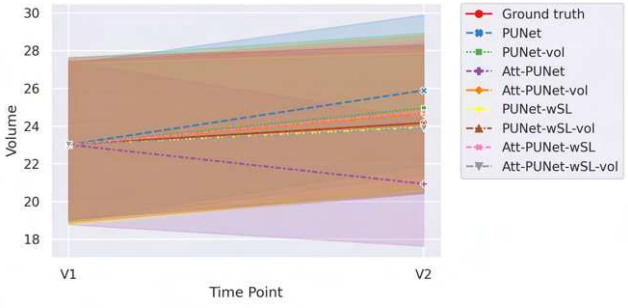

Figure 11: Average progression of WMH volume (*ml*) from V1 to V2 (1 year) for Ground truth and all tested models/configurations.

5.3. Volume based evaluation

WMH volume is an important clinical feature for clinical research and could be an important biomarker if available for clinical practice. Hence, we evaluated how well WMH volume at V2 (1 year later) can be estimated by using our proposed models. Table 6 shows the prediction accuracy of WMH volumetric progression (i.e., whether WMH volume will grow or shrink at V2 for each patient) calculated using Equations 10 and 11, the estimated volume interval (EVI) calculated using Equations 12, 13, and 14, and the volumetric error calculated using Equation 15. Figure 11 shows the average progression of WMH volume from V1 to V2 from the ground truth and all tested models/configurations. Figures 12 and 13 show volumetric progression of WMH (in *ml*) from V1 to V2 and its distribution for all patients in our dataset, and Figure 14 shows the WMH volume change in patients grouped by quintiles depending on their WMH volume at baseline.

As Table 6 shows, PUNet-wSL-vol performed better than the rest of the models producing either the best or second best results for almost all evaluation metrics except predicting growing WMH (GRW). Although there were more patients with net growing WMH than with net shrinking WMH in the dataset, thus hinting to a possible bias by the other models towards growing WMH, reduction in WMH volume was mainly observed in patients with high WMH volume (see Figure 14).

As Figure 11 shows, the average progression of WMH volume from V1 to V2 (in *ml*) was well estimated by PUNet-wSL-vol (i.e., brown dashed line representing PUNet-wSL-vol is coincident with the red line representing the ground truth). In general, as expected, models trained using volume loss (Equation 3, Section 4.4.2) (i.e., PUNet-vol (green line), Att-PUNet-vol (orange line), PUNet-wSL-vol (brown line), and Att-PUNet-wSL-vol (grey line)) produced more accurate progression of WMH volume from V1 to V2 than those which did not use volume loss during training. Of note, however, PUNet-wSL (yellow line) and Att-PUNet-wSL (pink line), had lines close to the red line of the ground truth. Overall, models jointly segmenting stroke lesions and WMH improved the estimation of future volume of WMH at V2 (see correspondence with average results sown in Table 6) .

Figures 12 and 13 illustrate that the WMH volumes at V2 estimated using PUNet-wSL-vol are more similar to the ground truth than the other models for most individual patients (Figure 12) and as a whole (Figure 13).

To further analyse the accuracy of the winner scheme in estimating the WMH volume change, we grouped the

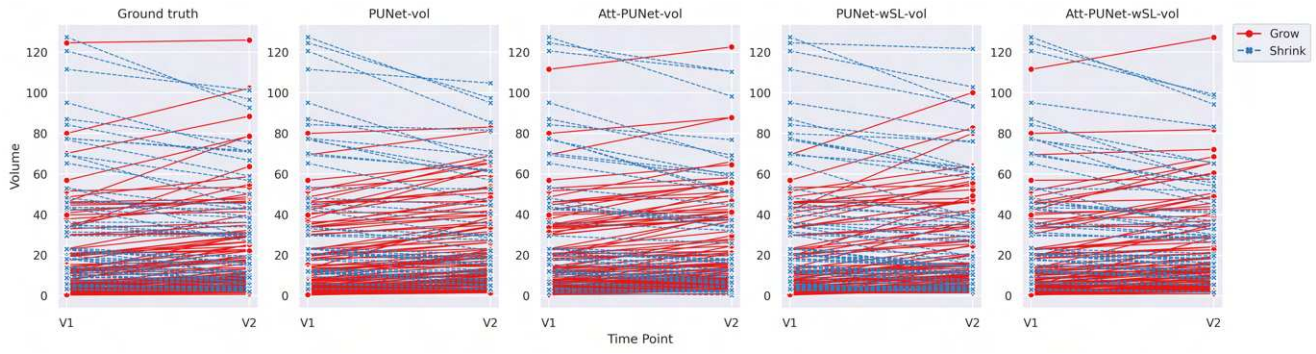

Figure 12: Volumetric progression of WMH (*ml*) from V1 to V2 (1 year apart) for each individual subject based on ground truth data (left) and future volume of WMH at V2 predicted by PUNet-vol, Att-PUNet-vol, PUNet-wSL-vol, and Att-PUNet-wSL-vol models.

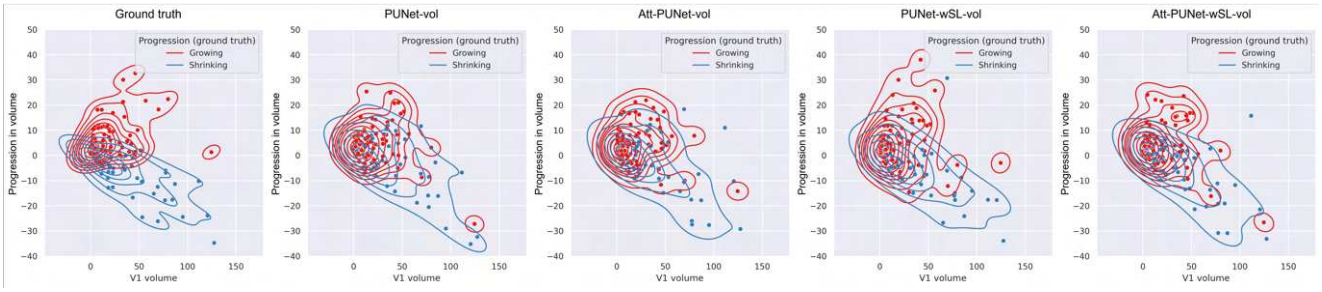

Figure 13: Distribution of volumetric progression of WMH at V2 (*ml*) based on WMH volume at V1 (*ml*) for each individual subject based on ground truth data (left) and predicted by PUNet-vol, Att-PUNet-vol, PUNet-wSL-vol, and Att-PUNet-wSL-vol models.

patients in quintiles according to their WMH volume at baseline and calculated the WMH change produced by the reference (i.e., ground truth) segmentation, and the PUNet model using volume loss with and without jointly segmenting the DEM and the stroke lesions. As can be appreciated from Figure 14, the scheme that jointly segmented the stroke lesions and the DEM of WMH change produced mean, median and a distribution of WMH volume change values across the sample more similar to those from the reference segmentation, than the scheme that only segmented the DEM of WMH change for all but the highest quintile.

We also divided the reference WMH segmentations into intense and less intense WMH as per (Valdés Hernández et al., 2015), and considered an 'extended' WMH volume

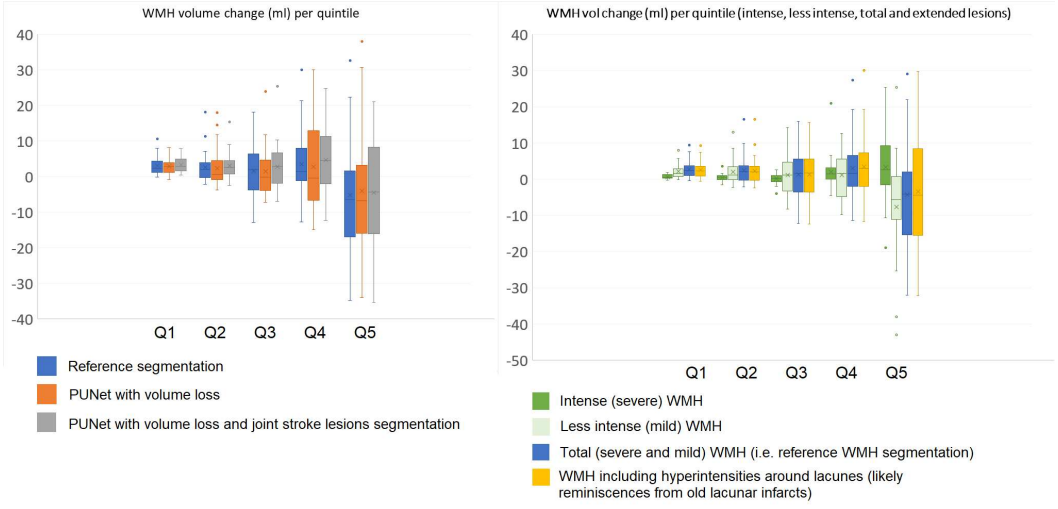

Figure 14: Volumetric WMH change in *ml* (vertical axes) for patients grouped by quintiles (horizontal axes) depending on their WMH volume at baseline.

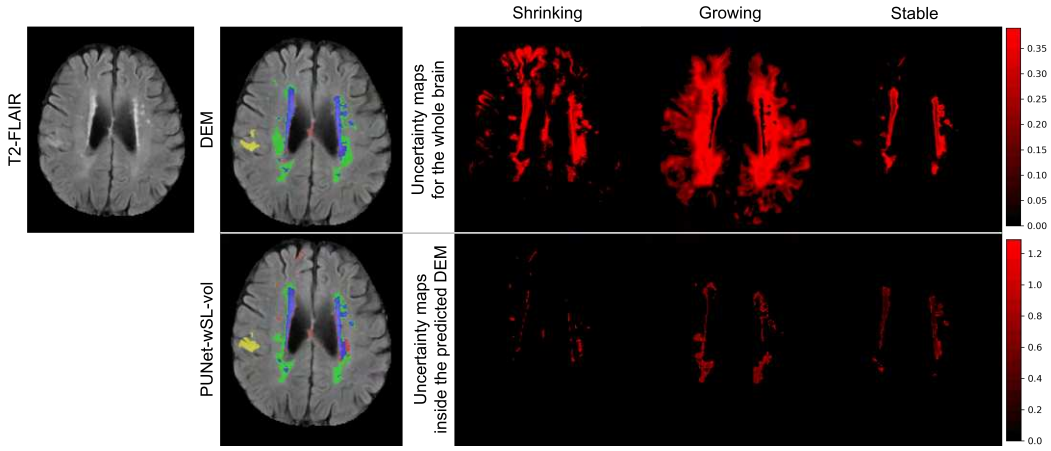

Figure 15: Uncertainty maps produced by PUNet-wSL-vol from subject MSSB212.

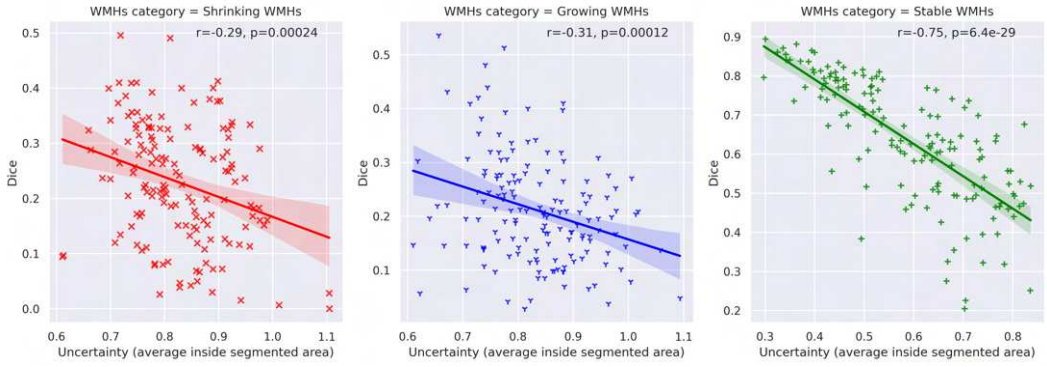

Figure 16: Correlation between the average of uncertainty values inside the predicted DEM and DSC values of predicted DEM produced by PUNet-wSL-vol for each DEM label.

that included the WMH surrounding lacunes, thought to be reminiscences of old small subcortical infarcts. As can be observed from Figure 14, the volume output from the scheme that jointly segmented the stroke lesions with the DEM of WMH change resulted strikingly similar to the one produced by this 'extended' WMH segmentation (see gray and yellow box plots in Figure 14), especially for patients in the highest quintile. Patients in this quintile exhibit a high burden of WMH coalescing with lacunes and previous strokes. Therefore, it is expected that not only AI schemes, but also experts would consider all hyperintensities as part of the white matter disease in absence of any other sequence or clinical information from this patient group. It can be also seen that the reference WMH change (i.e., blue box plot in the same figure) is mainly determined by the less intense WMH change (i.e., pale green box plot), therefore explaining the difficulty in obtaining accurate growth and shrinking spatial estimates and putting into question the accuracy in the spatial estimates of the ground truth segmentations given the degree of observer-dependent manual input they had.

5.4. Uncertainty quantification

As all configurations evaluated are based on the Probabilistic U-Net, uncertainty for each label in the DEM was quantified by predicting DEM for each subject multiple times. In this study 30 different DEM predictions were generated from 30 samples of z_{prior} from Prior Net for each input data/patient. From these 30 DEM predictions per patient data, uncertainty was calculated as the Cross Entropy (CE) between probability values from all DEM predictions and its average as written in Equation 9.

Figure 15 shows the uncertainty maps for all DEM labels produced by the model that generated the best DSC 'Average' value, PUNet-wSL-vol, for the whole brain and inside the predicted DEM for a patient. From the uncertainty maps for the whole brain, we can see that the uncertainties for shrinking and growing WMH encompass larger brain areas than for stable WMH. Some areas showing uncertainty in the 'Shrinking' label are incorrect (e.g. in the frontal cortex and in the septum), due mainly to hyperintense flow artefacts. Also, in the uncertainty maps inside the predicted DEM, the uncertainty values inside DEM labels of shrinking and growing WMH are higher than those inside stable WMH, a consistent finding from this evaluation.

Figure 16 shows that the uncertainty values inside the predicted DEM and DSC values produced by PUNet-wSL-vol are negatively correlated for each DEM label (i.e.,

shrinking, growing and stable WMH).The average uncertainty values inside each DEM label (i.e., the bottom figure in Figure 15) can be used to indicate the quality of predicted DEM. Higher DSC values of DEM labels can be predicted for lower uncertainty values inside the predicted DEM and vice versa.

6. Conclusion

This study proposed deep learning models based on the Probabilistic U-Net (Kohl et al., 2018) architecture trained incorporating stroke lesions information into the models and using volume loss as additional loss for improving the quality of predicted disease evolution map (DEM). Probabilistic U-Net was chosen as the baseline method because a preliminary study showed that it performed better than the U-Net in predicting DEM of WMH (Rachmadi et al., 2021). Based on our experiments, both proposed approaches improved the quality of predicted DEM of WMH in different ways.

We proposed three different approaches for incorporating stroke lesions information into Probabilistic U-Net models. These are **(1)** joint segmentation of DEM and stroke lesions (described in Section 3.2.1), **(2)** use of probabilistic maps of WMH change in relation to stroke lesions' locations (described in Section 3.2.2), and **(3)** combination of **(1)** and **(2)**. We proposed to incorporate stroke lesions information into deep learning models to predict WMH evolution because stroke is commonly associated with the evolution of WMH (Wardlaw et al., 2017). Based on results from various experiments, joint segmentation of DEM and stroke lesions (approach **(1)**) was the most effective approach to improve the quality of predicted DEM of WMH in all evaluations, while being also simpler and more straightforward than the other approaches evaluated in this study.

The introduction of a volume loss as an additional loss to the scheme substantially improved the quality in predicting the DEM of WMH in terms of volume (discussed in Section 5.3) and in terms of spatial agreement on 'Average' as DSC metric values showed (discussed in Section 5.1).

This study shows that 1) incorporating factors that have been commonly associated with WMH progression (i.e., stroke lesions information) is crucial to produce better prediction of DEM for WMH from brain MRI; 2) the best method for incorporating associated factors that can be extracted from the same data/image modality involves performing multi-task learning; and 3) in patients with vascular pathology, a multi-class segmentation of brain features resulting from symptomatic (i.e. stroke) and asymptomatic (i.e., WMH) vascular events generates better results consistent with clinical research. In this study, as stroke lesions appear on the same T2-FLAIR MRI sequence as WMH, we performed joint segmentation of DEM for WMH and stroke lesions. However, previous clinical studies have shown that there are other non-image risk factors and brain features that have been commonly associated with the progression and evolution of WMH, like age (van Dijk et al., 2008), ventricular enlargement (Breteler et al., 1994; Jochems et al., 2022b), and brain atrophy (Wardlaw et al., 2015). Thus, more (image and non-image) factors could be incorporated in future studies to further improve the quality of predicted DEM of WMH, although the best way to incorporate non-image factors to the prediction model remains to be found.

This study also has limitations to overcome in future works. The dataset was small in size impeding a quantitative in-depth analysis of the models' performance in different patient subgroups, e.g., patients stratified by age and sex, patients grouped by stroke subtype, etc. Thus, subgroup analyses were carried out visually and volumetrically, not spatially. By using only data from patients presenting to a clinic with a mild-to-moderate stroke, the generalisability of the proposed approach can be questioned. Therefore, further evaluation in a wider and more heterogeneous sample will be needed. The use of DSC in the evaluation needed the binarisation of the probabilistic outputs from the models. Limitations in the use of DSC have been recently acknowledged (Maier-Hein et al., 2022). However, it must be noted that ground truth segmentations are also binary, and observer-dependent. By using different quality control metrics in a comprehensive analysis we have overcome the limitations that the analysis of the spatial agreement using DSC poses. A probabilistic metric allowing spatial analyses of segmentation results is needed. Also, we used probabilistic maps of WMH change for strokes in the lentiform nucleus and centrum semiovale based on findings from a clinical study. However, the same clinical study specified that it was not possible to ascertain WMH evolution and distribution for patients with the stroke in other regions like thalami and midbrain or brain stem due to the limited sample of patients with infarcts in those regions. Incorporating findings for more powered studies would be necessary to conclude about the usefulness of incorporating attention maps to the AI schemes. Finally, various schemes for estimating uncertainty in segmentation/classification tasks have recently emerged (Liu et al., 2020; Sensoy et al., 2018), which would be worth exploring in the future for estimating WMH evolution.

CRediT authorship contribution statement

Muhammad Febrian Rachmadi: Conceptualization of this study, Methodology, Software, Experiments, Evaluations, Writing - Original draft preparation. **Maria del C. Valdés-Hernández:** Conceptualization of this study, Methodology, Clinical perspective and evaluation, Writing. **Stephen Makin:** Data curation. **Joanna Wardlaw:** Clinical perspective and evaluation, Data curation. **Henrik Skibbe:** Methodology, Evaluation, Writing.

Acknowledgements

Funds from JSPS (Kakenhi Grant-in-Aid for Research Activity Start-up, Project No. 20K23356) (MFR); RIKEN's Special Postdoctoral Researchers (SPDR) program (MFR); Row Fogo Charitable Trust (Grant No. BRO-D.FID3668413) (MCVH); Wellcome Trust (patient recruitment, scanning,

primary study Ref No. WT088134/Z/09/A); Fondation Leducq (Perivascular Spaces Transatlantic Network of Excellence); EU Horizon 2020 (SVDs@Target); and the MRC UK Dementia Research Institute at the University of Edinburgh (Wardlaw programme) are gratefully acknowledged. This research was also supported by the program for Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS) from the Japan Agency for Medical Research and Development AMED (JP15dm0207001). Library access provided by the Faculty of Computer Science, Universitas Indonesia is also gratefully acknowledged.

References

Breteler, M., van Amerongen, N.M., van Swieten, J.C., Claus, J.J., Grobbee, D.E., Van Gijn, J., Hofman, A., Van Harskamp, F., 1994. Cognitive correlates of ventricular enlargement and cerebral white matter lesions on magnetic resonance imaging. the rotterdam study. Stroke 25, 1109–1115.

Cai, M., Jacob, M.A., van Loenen, M.R., Bergkamp, M., Marques, J., Norris, D.G., Duering, M., Tuladhar, A.M., de Leeuw, F.E., 2022. Determinants and temporal dynamics of cerebral small vessel disease: 14-year follow-up. Stroke , 10–1161.

Chappell, F.M., del Carmen Valdés Hernández, M., Makin, S.D., Shuler, K., Sakka, E., Dennis, M.S., Armitage, P.A., Muñoz Maniega, S., Wardlaw, J.M., 2017. Sample size considerations for trials using cerebral white matter hyperintensity progression as an intermediate outcome at 1 year after mild stroke: Results of a prospective cohort study. Trials 18, 1–10. doi:10.1186/s13063-017-1825-7.

Chong, S.L., Liu, N., Barbier, S., Ong, M.E.H., 2015. Predictive modeling in pediatric traumatic brain injury using machine learning. BMC medical research methodology 15, 1–9.

Clayden, J.D., Maniega, S.M., Storkey, A.J., King, M.D., Bastin, M.E., Clark, C.A., 2011. Tractor: magnetic resonance imaging and tractography with r. Journal of Statistical Software 44, 1–18.

De Souza, F.S.H., Hojo-Souza, N.S., Dos Santos, E.B., Da Silva, C.M., Guidoni, D.L., 2021. Predicting the disease outcome in covid-19 positive patients through machine learning: a retrospective cohort study with brazilian data. Frontiers in Artificial Intelligence 4.

Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26, 297–302.

van Dijk, E.J., Prins, N.D., Vrooman, H.A., Hofman, A., Koudstaal, P.J., Breteler, M.M.B., 2008. Progression of cerebral small vessel disease in relation to risk factors and cognitive consequences: Rotterdam scan study. Stroke 39, 2712–2719. doi:10.1161/STROKEAHA.107.513176.

Godin, O., Tzourio, C., Maillard, P., Alpérovitch, A., Mazoyer, B., Dufouil, C., 2009. Apolipoprotein E genotype is related to progression of white matter lesion load. Stroke 40, 3186–3190.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: Advances in neural information processing systems, pp. 2672–2680.

Hüllermeier, E., Waegeman, W., 2021. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Machine Learning 110, 457–506. doi:doi.org/10.1007/s10994-021-05946-3.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.

Jochems, A.C., Arteaga, C., Chappell, F., Ritakari, T., Hooley, M., Doubal, F., Maniega, S.M., Wardlaw, J.M., 2022a. Longitudinal changes of white matter hyperintensities in sporadic small vessel disease: a systematic review and meta-analysis. Neurology 99, e2454–e2463.

Jochems, A.C., Maniega, S.M., Valdés Hernández, M.d.C., Barclay, G., Anblagan, D., Ballerini, L., Meijboom, R., Wiseman, S., Taylor, A.M., Corley, J., et al., 2022b. Contribution of white matter hyperintensities to ventricular enlargement in older adults. NeuroImage: Clinical 34, 103019.

Kohl, S., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J.R., Maier-Hein, K., Eslami, S.M.A., Jimenez Rezende, D., Ronneberger, O., 2018. A probabilistic u-net for segmentation of ambiguous images, in: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Curran Associates, Inc.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.

Liu, J., Lin, Z., Padhy, S., Tran, D., Bedrax Weiss, T., Lakshminarayanan, B., 2020. Simple and principled uncertainty estimation with deterministic deep learning via distance awareness. Advances in Neural Information Processing Systems 33, 7498–7512.

Luo, X., Jiaerken, Y., Yu, X., Huang, P., Qiu, T., Jia, Y., Li, K., Xu, X., Shen, Z., Guan, X., Zhou, J., Zhang, M., Adni, F.T.A.D.N.I., 2017. Associations between APOE genotype and cerebral small-vessel disease: a longitudinal study. Oncotarget 8, 44477–44489. doi:10.18632/oncotarget.17724.

Maier-Hein, L., Reinke, A., Christodoulou, E., Glocker, B., Godau, P., Isensee, F., Kleesiek, J., Kozubek, M., Reyes, M., Riegler, M.A., et al., 2022. Metrics reloaded: Pitfalls and recommendations for image analysis validation. arXiv preprint arXiv:2206.01653 .

Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y., 2018. Spectral normalization for generative adversarial networks. arXiv preprint arXiv:1802.05957 .

Moriya, Y., Kozaki, K., Nagai, K., Toba, K., 2009. Attenuation of brain white matter hyperintensities after cerebral infarction. American Journal of Neuroradiology 30, 3174. doi:10.3174/ajnr.A1340.

Nakagawa, T., Ishida, M., Naito, J., Nagai, A., Yamaguchi, S., Onoda, K., Initiative, A.D.N., 2020. Prediction of conversion to alzheimer's disease using deep survival analysis of mri images. Brain communications 2, fcaa057.

Nguyen, M., He, T., An, L., Alexander, D.C., Feng, J., Yeo, B.T., Initiative, A.D.N., et al., 2020. Predicting alzheimer's disease progression using deep recurrent neural networks. NeuroImage 222, 117203.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., 2018. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 .

Pease, M., Arefan, D., Barber, J., Yuh, E., Puccio, A., Hochberger, K., Nwachuku, E., Roy, S., Casillo, S., Temkin, N., et al., 2022. Outcome prediction in patients with severe traumatic brain injury using deep learning from head ct scans. Radiology .

Pellegrini, E., Ballerini, L., Valdés Hernández, M.d.C., Chappell, F.M., González-Castro, V., Anblagan, D., Danso, S., Muñoz-Maniega, S., Job, D., Pernet, C., et al., 2018. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 10, 519–535.

Pinto, M.F., Oliveira, H., Batista, S., Cruz, L., Pinto, M., Correia, I., Martins, P., Teixeira, C., 2020. Prediction of disease progression and outcomes in multiple sclerosis with machine learning. Scientific reports 10, 1–13.

Prins, N.D., Scheltens, P., 2015. White matter hyperintensities, cognitive impairment and dementia: an update. Nature reviews. Neurology 11, 157–65. doi:10.1038/nrneurol.2015.10.

Rachmadi, M.F., Valdés-Hernández, M.d.C., Makin, S., Wardlaw, J., Komura, T., 2020. Automatic spatial estimation of white matter hyperintensities evolution in brain mri using disease evolution predictor deep neural networks. Medical image analysis 63, 101712.

Rachmadi, M.F., Valdés-Hernández, M.d.C., Makin, S., Wardlaw, J.M., Komura, T., 2019. Predicting the evolution of white matter hyperintensities in brain mri using generative adversarial networks and irregularity map, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 146–154.

Rachmadi, M.F., Valdés-Hernández, M.d.C., Maulana, R., Wardlaw, J., Makin, S., Skibbe, H., 2021. Probabilistic deep learning with adversarial training and volume interval estimation-better ways to perform and

evaluate predictive models for white matter hyperintensities evolution, in: International Workshop on PRedictive Intelligence In MEdicine, Springer. pp. 168–180.

Rafael-Palou, X., Aubanell, A., Ceresa, M., Ribas, V., Piella, G., Ballester, M.A.G., 2022. Prediction of lung nodule progression with an uncertainty-aware hierarchical probabilistic network. Diagnostics 12, 2639.

Ramirez, J., McNeely, A.A., Berezuk, C., Gao, F., Black, S.E., 2016. Dynamic progression of white matter hyperintensities in Alzheimer's disease and normal aging: Results from the Sunnybrook dementia study. Frontiers in Aging Neuroscience 8, 1–9. doi:10.3389/fnagi.2016.00062.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

Sachdev, P., Wen, W., Chen, X., Brodaty, H., 2007. Progression of white matter hyperintensities in elderly individuals over 3 years. Neurology 68, 214–222. doi:10.1212/01.wnl.0000251302.55202.73.

Sauty, B., Durrleman, S., 2022. Progression models for imaging data with longitudinal variational auto encoders, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 3–13.

Schmidt, H., Zeginigg, M., Wiltgen, M., Freudenberger, P., Petrovic, K., Cavalieri, M., Gider, P., Enzinger, C., Fornage, M., Debette, S., Rotter, J.I., Ikram, M.A., Launer, L.J., Schmidt, R., 2011. Genetic variants of the NOTCH3 gene in the elderly and magnetic resonance imaging correlates of age-related cerebral small vessel disease. Brain 134, 3384–3397. doi:10.1093/brain/awr252.

Schmidt, R., Enzinger, C., Ropele, S., Schmidt, H., Fazekas, F., 2003. Progression of cerebral white matter lesions: 6-Year results of the Austrian Stroke Prevention Study. Lancet 361, 2046–2048. doi:10.1016/S0140-6736(03)13616-1.

Schmidt, R., Fazekas, F., Enzinger, C., Ropele, S., Kapeller, P., Schmidt, H., 2002. Risk factors and progression of small vessel disease-related cerebral abnormalities, in: Ageing and Dementia Current and Future Concepts. Springer, pp. 47–52.

Sensoy, M., Kaplan, L., Kandemir, M., 2018. Evidential deep learning to quantify classification uncertainty. Advances in neural information processing systems 31.

Valdés Hernández, M.d.C., 2021. Human brain atlases across the lifespan – anatomical segmentations, 1990-2016 [dataset]. University of Edinburgh. Centre for Clinical Brain Sciences. URL: https://datashare.ed.ac.uk/handle/10283/3890.

Valdés Hernández, M.d.C., Armitage, P.A., Thrippleton, M.J., Chappell, F., Sandeman, E., Muñoz Maniega, S., Shuler, K., Wardlaw, J.M., 2015. Rationale, design and methodology of the image analysis protocol for studies of patients with cerebral small vessel disease and mild stroke. Brain and behavior 5, e00415. doi:10.1002/brb3.415.

Valdés Hernández, M.d.C., Grimsley-Moore, T., Sakka, E., Thrippleton, M.J., Chappell, F.M., Armitage, P.A., Makin, S., Wardlaw, J.M., et al., 2021. White matter hyperintensities evolution patterns 1 year post-lacunar stroke and their association with post-stroke cognition, 2009-2013 [dataset]. University of Edinburgh. Centre for Clinical Brain Sciences. URL: https://datashare.ed.ac.uk/handle/10283/3934.

Valdés Hernández, M.d.C., Piper, R.J., Wang, X., Deary, I.J., Wardlaw, J.M., 2013. Towards the automatic computational assessment of enlarged perivascular spaces on brain magnetic resonance images: a systematic review. Journal of Magnetic Resonance Imaging 38, 774–785.

Valdés Hernández, M.d.C., Grimsley-Moore, T., Chappell, F.M., Thrippleton, M.J., Armitage, P.A., Sakka, E., Makin, S., Wardlaw, J.M., 2021. Post-stroke cognition at 1 and 3 years is influenced by the location of white matter hyperintensities in patients with lacunar stroke. Frontiers in Neurology 12. URL: https://www.frontiersin.org/articles/10.3389/fneur.2021.634460, doi:10.3389/fneur.2021.634460.

Wang, X., Hernández, M.C.V., Doubal, F., Chappell, F.M., Wardlaw, J.M., 2012. How much do focal infarcts distort white matter lesions and global cerebral atrophy measures? Cerebrovascular Diseases 34, 336–342.

Wardlaw, J.M., Chappell, F.M., Valdés Hernández, M.D.C., Makin, S.D., Staals, J., Shuler, K., Thrippleton, M.J., Armitage, P.A., Munõz-Maniega, S., Heye, A.K., Sakka, E., Dennis, M.S., 2017. White matter hyperintensity reduction and outcomes after minor stroke. Neurology 89, 1003–1010. doi:10.1212/WNL.0000000000004328.

Wardlaw, J.M., Smith, E.E., Biessels, G.J., Cordonnier, C., Fazekas, F., Frayne, R., Lindley, R.I., O'Brien, J.T., Barkhof, F., Benavente, O.R., Black, S.E., Brayne, C., Breteler, M., Chabriat, H., Decarli, C., de Leeuw, F.E., Doubal, F., Duering, M., Fox, N.C., Greenberg, S., Hachinski, V., Kilimann, I., Mok, V., Oostenbrugge, R.v., Pantoni, L., Speck, O., Stephan, B.C.M., Teipel, S., Viswanathan, A., Werring, D., Chen, C., Smith, C., van Buchem, M., Norrving, B., Gorelick, P.B., Dichgans, M., STandards for ReportIng Vascular changes on nEuroimaging (STRIVE v1), 2013. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. The Lancet. Neurology 12, 822–38. doi:10.1016/S1474-4422(13)70124-8.

Wardlaw, J.M., Valdés Hernández, M.C., Muñoz-Maniega, S., 2015. What are white matter hyperintensities made of? relevance to vascular cognitive impairment. Journal of the American Heart Association 4, e001140.

Zhang, T., Liao, Q., Zhang, D., Zhang, C., Yan, J., Ngetich, R., Zhang, J., Jin, Z., Li, L., 2021. Predicting mci to ad conversation using integrated smri and rs-fmri: machine learning and graph theory approach. Frontiers in Aging Neuroscience , 429.